Reviewer’s report

Title: Characteristic distribution of the total and individual item scores on the Kessler Screening Scale for Psychological Distress (K6) in US adults

Version: 0 Date: 16 May 2017

Reviewer: Rob Saunders

Reviewer's report:

This is an interesting paper which explores the distribution of depressive symptom scores in a US population using national survey data. The research follows on from analyses of similar data from other sources by the authors, and is informative. I have provided some comments below which should be addressed to aid the interpretation and reach of this study, but also feel the authors need to improve the underpinning of this work - it has a lot of potential value but is underplayed at the moment.

Two general points first, before comments by section:

* The phrase "lower end of the distribution" is used nearly 30 times in the manuscript which is a bit too much. Please change the wording periodically to aid readability as it seem repetitive at the moment - mixing in phrases like 'lowest symptom counts', 'individuals with low depression scores'.

* Although the focus is on the 'lower end' it appears from the results and as you suggest within the paper that the 'higher' end also has lower frequencies compared to those predicted from an exponential pattern. However, this seems largely ignored in your discussion with the focus more on your 'lower-end' theory. I feel a bit more balance is needed, as there is a danger readers might think you are ignoring the higher end and focusing on your lower-end theory instead.

Title

*Typo - please correct (chatacterisite patten).

* I also recommend it is changed slightly, as currently it isn't that clear. The phrase 'sum of the depressive symptom item scores' doesn't seem right - wouldn't changing to 'distribution of total and individual item scores' be more clear considering the contents of the paper?
Abstract:

* Start with "The distributional pattern…"

* Replace "lower end of the distribution" with 'lower scorers', or similar.

Introduction.

* The very first sentence is copied directly from your 2015 and 2016 papers, please amend.

* Suggest re-phrasing sentence "because depressive symptomology…" - diagnosis is arrived at in relation to the amount of symptoms as well as their severity and chronicity. Explain why understanding the distribution of depressive symptoms in the general population is important? Your description is limited but it is the justification of this study.

* Suggest re-writing first sentence, second paragraph. Increasing sample size reduces uncertainly that findings reflect the population. There's an assumption a larger sample more closely approximates the population distribution, but it isn't guaranteed.

* "exponential pattern, except for the lower end of the distribution" - this needs to be defined for the general reader of this journal, who may not be as comfortable with the concepts used in this paper. This concept is fundamental to the interpretation of the findings later.

* I would suggest that you change the depression of the K6 items as saying "anxiety" is a 'depressive symptom' might cause confusion (and clinical disagreement). Changing to 'nervousness' or 'worry' might be safer.

* Rephrase the sentence "Furthermore, we investigated the total scores of the PHQ" to make it clear that it's an upcoming study and place the reference at the end of the first sentence rather than end of the paragraph. No issue with describing finding in this part though, as it's relevant.

* With the following paragraph ("taken together,…"), I think this is an important part of the introduction but it isn't strong enough. It links to a part about the assumption of normal distributions and psychometric methods (i.e. item response theory) you make in the discussion. This justifies the value of this study.
Method

* I wonder why the section starting "Of the RDD respondents who reported…" to "….solicit their participation in the survey" is of value to this study? I think you could drop and spend the words elsewhere. I think you could cut the dataset section even further as surely it's presented in a lot of detail elsewhere.

*Analysis procedure - please provide percentage of individuals excluded from the sample.

*As so much analysis has been conducted as part of this study, the 'analysis procedure' section needs to be clear, and would benefit from a bit more detail. The part about "non-exponential distribution at the lowest end of the scores" could be more clearly defined, and why calculating ratios helps this.

*Linked to above, why was a log-normal scale and regression analysis used? This may not be obvious to all readers, who may not be knowledgeable about methods but it is important to understand why you have done this for the interpretation of the results.

Results

* Figure 1a/b - I couldn't see item 4 in this graph at all. It might be hiding behind another line but at present it appears you've missed it. Please comment on it at least, as looks like an omission.

* Page 10 - "…after which they show a decreasing pattern" - I'd argue they were all decreasing before the crossover anyway, so please say a little more about this. It's an interesting finding.

* Page 11 start - change 'inevitably' to 'as expected'.

* "total scores of the 6 items showed linear patterns…" - there is also quite a bit of fluctuation as well, especially at the lower end which deserves comment. Appreciate it's likely due to lower numbers at this level of symptoms. Fluctuation needs a bit a comment in the discussion as well.
*Typo - Ration instead is ratio.

Discussion

*Major point with the discussion and first paragraph - although you have focused on the 'lower end of the distribution' throughout the paper, the higher end of the distribution is equally interesting but hasn't much mention in the discussion at all. In not every set of graphs you present is your statement correct and therefore a bit more balance to your argument is needed. There's a danger that the reader might interpret this as you ignoring the higher end of the distribution just to fit with your 'lower end' hypothesis.

* Conclusion - whereas the rest of the discussion is really strong, the conclusion looks rather weak at the moment in comparison which is a shame as this is the take home message. I suggest it's re-written to improve. The first sentence is the same statement repeated. The next sentence doesn't have the desired effect as it's not clear why this information in the general population impacts clinical practice, and this isn't mentioned elsewhere adequately in this manuscript - but I suggest it is to improve reach.

Figure legends

*Two things with these:

1). You have written 'rate' rather than 'ratio' for all legends/graphs which appears wrong.

2) I recommend in the legends you specify which items are in the low/medium/high groups for each set of graphs.
Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

Yes

Does the work include the necessary controls?
If not, please specify which controls are required in your comments to the authors.

Unable to assess

Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.

No

Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I am able to assess the statistics

Quality of written English
Please indicate the quality of language in the manuscript:

Acceptable

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?
If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

No interests to declare.

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal